

GUNADARMA UNIVERSITY

MASTER DEGREE PROGRAM



A/B Testing Mekargo.id Website of Mekar PT Sampoerna Wirausaha Using Bayesian Inference

Written by:

Name : Alvian Dwi Kurnianto, S.Kom.
Student Id : 92216007
Supervisor : Dr. Setia Wirawan

Submitted as a Partial Fulllment of the Requirements for
Magister Degree of Information System Management
Concentrating on Information System Software

Depok

2018

PAGE OF APPROVAL

Title : "A/B Testing Mekargo.id Website of Mekar
PT Sampoerna Wirausaha Using Bayesian
Inference"
Name : Alvian Dwi Kurnianto, S.Kom.
NIM : 92216007
Date of Graduation : May, 12 2017

Approved by:

Board of Advisors

.....

Dr. Setia Wirawan

(Chair Person)

.....

Prof. Dr. Yuhara Sukra., M.Sc.

(Member)

.....

Dr. Tb. Maulana Kusuma, S.Kom., MEngSc.

(Director)

ABSTRACT

Alvian Dwi Kurnianto, S.Kom.. 92216007

**"A/B TESTING MEKARGO.ID WEBSITE OF MEKAR PT SAMPOERNA WIRAUSAHA
USING BAYESIAN INFERENCE".**

Master Thesis, Information System Management, Information System Software, Gunadarma University, 2018.

Keyword : Music Player, Application, Android, Soundcloud, API, Responsive web, Mobile Application, Music.

(xiii+ 54)

References (2013-2017)

ACKNOWLEDGEMENTS

Praise and thanks raised to Allah SWT for all blessing and also utterance and good wishes sent to the prophet of Muhammad SAW so, with all the grants, author have finished this thesis, which is titled "A/B Testing Mekargo.id Website of Mekar PT Sampoerna Wirausaha Using Bayesian Inference". This undergraduate thesis is intended to complete the requirement to finish my study in the Informatics Engineering Department, Gunadarma University. This writing would not have been possible without the support and encouragement from people around me. Therefore, I would like to say thank to:

.

Depok, February 2017

Alvian Dwi Kurnianto, S.Kom.

Contents

Abstract	iii
Acknowledgements	iv
List of Figures	vii
List of Tables	viii
1 INTRODUCTION	1
1.1 Background	1
1.2 Formulation of the Problem	2
1.3 Scope of Problem	2
1.4 Research Purposes	3
1.5 Research Benefits	3
2 LITERATURE REVIEW	4
2.1 Python	4
2.2 Numpy	5
2.3 Matplotlib	5
2.4 Scipy	5
2.5 Ipython	6
2.6 Jupyter	6
2.7 Bayesian Inference	7
2.8 Pymc	7
2.9 Google Analytics	7
2.10 A/B Testing	7
2.11 Previous Researches	7

3 RESEARCH METHODOLOGY	8
3.1 Research Object	8
3.2 Research Methodology	8
3.2.1 Identify Problems	9
3.2.2 Define Website Measurement	9
3.2.3 Develop a Hypothesis	9
3.2.4 Developing and Testing Page Variants	10
3.2.5 Analyze Test Results	10
4 RESULTS AND DISCUSSION	11
4.1 Identify Problems	11
4.2 Define Website Measurement	11
4.2.1 Business Objective	11
4.2.2 Website Goal	12
4.2.3 Key Performance Metric	12
4.2.4 Target Metric	12
4.3 Develop a Hypothesis	12
4.4 Developing and Testing Page Variants	12
4.4.1 Developing Page Variants	12
4.4.2 Testing Page Variants	15
4.5 Analyze Test Results	16
5 CONCLUSION AND SUGGESTIONS	22
5.1 Conclusion	22
5.2 Suggestions	22
Bibliography	23
Appendix	25
Curriculum Vitae	26

List of Figures

3.1	Research	9
4.1	Graphic google analytics mekargo.id	15
4.2	Table google analytics mekargo.id	16
4.3	Import python library	16
4.4	Initialize sample	17
4.5	Calculate probability	17
4.6	Bernoulli observation	17
4.7	Sum and mean of bernoulli observation	18
4.8	Pymc probabilistic model	19
4.9	Pymc probabilistic model results	19
4.10	Posteriror distribution P_A	20
4.11	Posteriror distribution P_B	20
4.12	Posteriror distribution P_C	21
4.13	Posteriror distribution delta A-B	21

List of Tables

2.1	Previous research	7
4.1	Data fields changes	13
4.2	Url definition	13

Chapter 1

INTRODUCTION

1.1 Background

Usaha Kecil Menengah (UKM) memiliki peran penting dalam mendorong pertumbuhan perekonomian Indonesia. Dengan adanya sektor UKM, pengangguran akibat angkatan kerja yang tidak terserap dalam dunia kerja menjadi berkurang. Sektor UKM pun telah terbukti menjadi pilar perekonomian yang tangguh. Kontribusi sektor UKM dalam menentukan Produk Domestik Bruto (PDB) dan sektor penghasil devisa negara juga tak perlu diragukan lagi. Saat ini, UKM telah dijadikan agenda utama pembangunan ekonomi Indonesia (Kemenkeu, 2015).

Masalah mendasar UKM yang paling menonjol menyangkut menyediakan pembiayaan usaha alias modal usaha. Kebutuhan modal sangat terasa pada saat seseorang ingin memulai usaha baru. Alhasil, biasanya bila motivasinya kuat, seseorang akan tetap memulai usaha dengan modal seadanya. Pada usaha yang sudah berjalan, modal tetap menjadi kendala lanjutan untuk berkembang. Masalah yang menghadang UKM menyangkut kemampuan akses pembiayaan, akses pasar dan pemasaran, tata kelola manajemen usaha kecil serta akses informasi. Kesulitan UKM mengakses sumber-sumber modal karena keterbatasan informasi dan kemampuan menembus sumber modal tersebut. Padahal pilihan sumber modal sangat banyak dan beragam. Lembaga keuangan bank adalah sumber modal terbesar yang dapat dimanfaatkan oleh pelaku usaha kecil. Namun untuk bermitra dengan bank, usaha kecil dituntut menyajikan proposal usaha yang feasible atau layak usaha dan menguntungkan.

Mekar (PT Sampoerna Wirausaha) didirikan pada 2010 dan dimiliki secara penuh oleh Putera Sampoerna Foundation, Mekar bertujuan untuk mempermudah akses finansial bagi para pelaku UKM Indonesia sehingga memiliki dampak sosial dan ekonomi yang positif di Indonesia (Mekar, 2017). Mekar memberikan solusi dalam membantu Pelaku Usaha Mikro Kecil dan Menengah (UMKM) dan Konsumen lainnya untuk mendapatkan Akses Layanan Keuangan atau Permodalan di Indonesia. Beberapa produk yang dimiliki Mekar diantaranya website utama Mekar.id, chat bot dan website Mekar Go. Website Mekar Go dengan alamat web <https://mekargo.id/> merupakan website untuk menghimpun data UKM dan karyawan yang ingin mengajukan pinjaman ke MEKAR. Data yang telah dihimpun pada versi web saat ini menunjukkan konversi peminjam yang melengkapi pengisian data dari awal sampai akhir kurang dari 10%. Presentase konversi ini menurut tim marketing terlalu kecil dan perlu ditingkatkan.

Untuk meningkatkan presentase konversi pelengkapan data, tim marketing mengusulkan perubahan urutan pengisian data. Perubahan urutan pengisian data dibuat menjadi dua versi baru yang selanjutnya disebut sebagai versi B dan versi C, dan versi saat ini disebut sebagai versi A. Ketiga versi tersebut akan dilakukan A/B testing untuk mengetahui versi manakah yang menghasilkan presentasi konversi pelengkapan data terbanyak. Pengumpulan data dibantu dengan tools Google Analytics. Setelah data yang diinginkan terpenuhi, dilakukan proses analisis hasil data dengan metode bayessian inference dan tools pymc3.

1.2 Formulation of the Problem

Adapun masalah yang dibahas pada penelitian ini adalah bagaimana metode yang dilakukan untuk melakukan a/b testing website mekargo.id pada PT Mekar Sampoerna Wirausaha, bagaimana cara menganalisis hasil a/b testing dengan metode bayesian inference dan tools pymc3, dan varian manakah dari A/B testing yang menjadi varian terbaik.

1.3 Scope of Problem

Adapun batasan masalah dari penelitian ini adalah sebagai berikut:

1. Objek penelitian adalah website mekargo.id
2. Penelitian dilakukan di PT Mekar Sampoerna Wirausaha
3. Waktu penelitian antara agustus 2017 sampai januari 2018
4. Penelitian menerapkan teknik a/b testing pada website mekargo.id dengan bantuan tools Google Analytics
5. Data yang dihasilkan dari A/B testing dilakukan analisis dengan metode bayesian inference menggunakan tools library pymc3, numpy, figsize, matplotlib, scipy, jupyter dan IPython dengan bahasa pemrograman python
6. Fitur website mekargo.id yang dibahas adalah fitur pengisian data pinjaman.
7. Fitur website mekargo.id yang tidak dibahas adalah fitur static page yang berisi halaman home, about us, terms and conditions, privacy policy dan contact us.

1.4 Research Purposes

Tujuan dari penelitian ini adalah melakukan a/b testing website mekargo.id pada PT Mekar Sampoerna Wirausaha, melakukan analisis hasil a/b testing dengan metode bayesian inference dengan pymc3, mengetahui variant pada A/B testing yang menjadi varian terbaik.

1.5 Research Benefits

Manfaat yang didapat dari penelitian ini adalah:

1. Bagi peneliti, menjadi panduan atau referensi penelitian di bidang sistem informasi pada penelitian selanjutnya.
2. Bagi peneliti lain, sebagai sumber literatur untuk penelitian yang sejenis.
3. Bagi perusahaan yang diteliti, memberikan saran perbaikan dengan terpilihnya varian a/b testing yang terbaik.

Chapter 2

LITERATURE REVIEW

2.1 Python

Python is a good general-purpose, high-level language. Its design makes it very readable, which is more important than it sounds. Every computer program is written only once, but read and revised many times, often by many people. Being readable also makes it easier to learn and remember, hence more writeable. Compared with other popular languages, Python has a gentle learning curve that makes you productive sooner, yet it has depths that you can explore as you gain expertise(Lubanovic, 2014).

People use python because have these several advantages: software quality, developer productivity, program portability, support libraries, component integration, and enjoyment(Lutz, 2013). Python is a multi-platform, general-purpose programming language that can run on Windows, Linux/Unix, and Mac OS X, and has been ported to Java and .NET virtual machines as well. It has a powerful standard library. In addition, it has many libraries for data analysis such as Numpy, Pandas, Matplotlib, PyMongo, and Scikit (Vo et al., 2015).

Python in this research is used as probabilistic programming tools for bayesian method to build probabilistic models. Probabilistic programming offers an effective way to build complex models and allows us to focus more on model design, evaluation, and interpretation, and less on mathematical or computational details (Martin, 2016).

2.2 Numpy

Numpy is one of the fundamental packages used for scientific computing in Python. Numpy is used as an efficient multidimensional container of generic data. Arbitrary data types can be defined and integrated with a wide variety of databases (Vo et al., 2015). The NumPy package, which comprises the NumPy array as well as a set of accompanying mathematical functions, has found wide-spread adoption in academia, national laboratories, and industry, with applications ranging from gaming to space exploration (Van Der Walt et al., 2016).

2.3 Matplotlib

Matplotlib is a python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms. Matplotlib can be used in python scripts, the python and ipython shell, web application servers, and six graphical user interface toolkits (Hunter et al., 2014). Matplotlib provides both a very quick way to visualize data from Python and publication-quality figures in many formats: line plots, contour plots, scatter plots, and Basemap plots. It comes with a set of default settings, but allows customization of all kinds of properties. However, we can easily create our chart with the defaults of almost every property in Matplotlib (Vo et al., 2015).

2.4 Scipy

SciPy (pronounced “Sigh Pie”) is a Python-based ecosystem of open-source software for mathematics, science, and engineering. SciPy ecosystem includes general and specialised tools for data management and computation, productive experimentation and high-performance computing. The SciPy library, a collection of numerical algorithms and domain-specific toolboxes, including signal processing, optimization, statistics and much more (Jones et al., 2014).

2.5 IPython

IPython is an interactive browser-based environment where you can combine code execution, text, mathematics, plots, and rich media into a single document. Originally designed for use as an electronic lab notebook for computational science, it is increasingly being used in teaching as well, and a rich ecosystem of open source plugins and extensions for teaching is growing around it (Wilson et al., 2014). IPython has provided terminal-based tools for interactive computing in Python since 2001. The notebook document format and multi-process architecture introduced in 2011 have expanded the applicable scope of IPython into teaching, presenting, and sharing computational work, in addition to interactive exploration (Ragan-Kelley et al., 2014).

2.6 Jupyter

Jupyter Notebook, provides a tool to create and share web pages with text, charts, and Python code in a special format. Often, the notebooks are used as an educational tool, or to demonstrate Python software. We can import or export notebooks either from plain Python code or from the special notebook format. The notebooks can be run locally, or we can make them available online by running a dedicated notebook server (Fandango and Idris, 2017).

Jupyter Notebook is accessed through a modern web browser. This makes it practical to use the same interface running locally like a desktop application, or running on a remote server. In the latter case, the only software the user needs locally is a web browser; so, for instance, a teacher can set up the software on a server and easily give students access. The notebook files it creates are a simple, documented JSON format, with the extension ‘.ipynb’. It is simple to write other software tools which access and manipulate these files (Kluyver et al., 2016).

2.7 Bayesian Inference

2.8 Pymc

2.9 Google Analytics

2.10 A/B Testing

2.11 Previous Researches

Penelitian sebelumnya

Table 2.1: Previous research

No	Journal	Authors	Strength	Weakness

Chapter 3

RESEARCH METHODOLOGY

Menjelaskan cara pelaksanaan kegiatan penelitian yang meliputi obyek penelitian, tempat dan waktu penelitian dilakukan, dan metode penelitian yang terdiri dari tahap-tahap penelitian dari awal sampai akhir berikut konsep / metodologi / algoritma yang dipakai dan bahasa pemrograman atau software tool yang digunakan.

3.1 Research Object

Waktu penelitian dilakukan antara bulan September 2017 sampai Januari 2018. Objek pada penelitian ini adalah website mekargo.id PT Mekar Investama Sampoerna. Fitur yang diteliti adalah fitur pendaftaran bagi pem-injam. Data yang digunakan adalah alamat url yang dikunjungi pendaftar untuk masing-masing versi a/b testing. Pengumpulan data dilakukan dengan tools google analytics. Data yang didapatkan selanjutnya akan dianalisis dengan metode bayesian inference dan tools pymc3.

3.2 Research Methodology

diagram 4 langkah besar, tulis 4.1.1-4.1.4.

Metode penelitian yang dilakukan meliputi: identifying problems, develop a hypothesis, developing and testing page variants, dan analyze test results. Gambaran langkah metode penelitian yang dilakukan dapat diilustrasikan oleh figure 3.1.

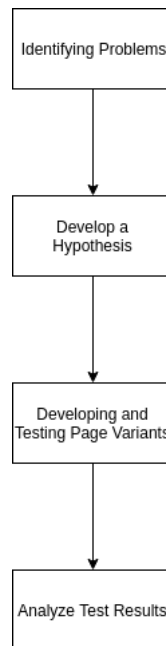


Figure 3.1: Research

3.2.1 Identify Problems

Masalah yang diidentifikasi adalah yang berhubungan dengan website yang akan dilakukan a/b testing. Masalah yang diangkat biasanya berhubungan dengan bisnis, marketing, atau conversion rate. Tergantung pada tujuan apa yang ingin dicapai pada pembuatan a/b testing pada website tersebut.

3.2.2 Define Website Measurement

Website measurement terdiri dari penjelasan poin indikator bisnis yang terdiri dari business objective, website goal, key performance metric, dan target metric.

Business objective adalah bla bal bla. Website goal adalah bla bla. Key performance metric adalah bla bla. Target metric adalah bla bla bla.

3.2.3 Develop a Hypothesis

lihat kata2 dari website

Asumsi awal yang menjadi penyebab masalah pada point 3.3.1 disertai dengan pengajuan solusi yang akan diterapkan. Menyebutkan hasil akhir dari solusi yang diajukan disertai alasannya.

3.2.4 Developing and Testing Page Variants

develop, menjelaskan tool analytic, penjelasan proses percobaan.

3.2.5 Analyze Test Results

menjelaskan konsep, metode, algoritma, dan tools program di pymc jupyter

Chapter 4

RESULTS AND DISCUSSION

Bab ini menjelaskan kegiatan hasil penelitian yang diperoleh seperti variabel yang diteliti, hasil analisis dan pengujian hipotesis dan pemrogramannya.

Penyajian dapat dibantu dengan menggunakan tabel, gambar, diagram atau grafik yang diuraikan secara jelas dengan memasukkan hasil penafsiran, analisis deskripsi dan kesimpulan pengujian hipotesis (Bila ada). Di dalam pembahasan juga dapat mengemukakan argumentasi dengan mengacu ke sumber pustaka yang relevan.

Adapun hasil dan pembahasan dari penelitian ini adalah sebagai berikut:

4.1 Identify Problems

jelaskan versi sebelum ab testing seperti apa, kondisinya bagaimana, yang dihasilkan, dan masalahnya apa, yang timbul dan penyebab kira kira

presentase pendaftar yang menyelesaikan pengisian data masih belum mencapai presentase minimal yang ditetapkan.

4.2 Define Website Measurement

4.2.1 Business Objective

Menambah jumlah peminjam yang akan diberikan pinjaman peer-to-peer lending.

To test which site is the highest successfully completed Loan Access and Survey submitted?

4.2.2 Website Goal

Meningkatkan jumlah peminjam yang mendaftar, meningkatkan presentase penyelesaian pengisian form pendaftaran oleh peminjam. Macro Goal: visitors completed submitting on loan access; Micro Goal: visitors completed submitting survey;

4.2.3 Key Performance Metric

jumlah pendaftar per bulan, presentase penyelesaian pengisian form oleh peminjam saat pengisian form pendaftaran.

Traffic Level, High and Low Traffic during the test period.
conversion level

4.2.4 Target Metric

500 pendaftar selama sebulan, presentase penyelesaian data di atas 10%.

4.3 Develop a Hypothesis

Presentase penyelesaian pengisian form pendaftaran dipengaruhi oleh urutan pengisian form. Apabila diubah akan meningkatkan presentasinya. Solusi yang diajukan adalah melakukan pengujian A/B testing terhadap dua versi baru disertai dengan penyederhanaan dan penyingkatan pengisian data. Asumsi awal, salah satu dari dua versi baru ini akan lebih baik daripada versi yang sebelumnya.

4.4 Developing and Testing Page Variants

4.4.1 Developing Page Variants

menjelaskan 3 versi, satu control 2 varian, dilihatin bentuk formnya.

perubahan yang dilakukan: Remove unnecessary data fields and some data fields are changed to optional

Table 4.1: Data fields changes

Part	Fields Name	Remove	Optional	New
Personal Detail Information	Alamat Usaha / Tempat Bekerja (Provinsi, Kota, Kecamatan, Kelurahan)	v		
	Bidang Jenis Usaha / Jenis Pekerjaan	v		
	Lama Usaha / Lama Bekerja	v		
	Jumlah Karyawan	v		
	Foto Tempat Usaha	v		
	Status Karyawan	v		
	Email		v	
	Foto KTP		v	
	Tanggal Lahir			v
Jaminan	Foto Tanah + Bangunan		v	
	Foto Kendaraan Bermotor		v	

Untuk membedakan versi a/b testing, digunakan suffik satu huruf alfabet (a atau b atau c) sesuai dengan jenis varian.

Table 4.2: Url definition

Step	Site A (/ukm/a/)	Site B (/ukm/b/)	Site C (/ukm/c/)
Survey Page	/ukm/a/survei	/ukm/b/survei	/ukm/c/survei
Detail Needs Page	/ukm/a/detil-survei	/ukm/b/detil-survei	/ukm/c/detil-survei
Personal Information Page	/ukm/a/data-diri	/ukm/b/data-diri	/ukm/c/data-diri

Step	Site A (/ukm/a/)	Site B (/ukm/b/)	Site C (/ukm/c/)
Thanks Page	/ukm/a/terima-kasih	/ukm/b/terima-kasih	/ukm/c/terima-kasih
Detail Loan Access Page	/ukm/a/ pinjaman	/ukm/b/ pinjaman	/ukm/c/ pinjaman
Ekstra Pesonal Detail Information Page	/ukm/a/ tambahan-data- diri	-	-
Congrats Page	/ukm/a/selamat	/ukm/b/selamat	/ukm/c/selamat

jelaskan dan sebutkan setiap alur url a/b/c (bisa dalam bentuk chart)

a Survey Page,

b Survey Page, Personal Information Page

c Survey Page,

kasih reference ke lampiran gambar

pembagian traffic round robbin

```

1 class RoundRobin(models.Model):
2     flow_type = models.CharField(max_length=1, null=True)
3     partner_slug = models.SlugField(null=True)
4     def next(self):
5         latest = self.flow_type
6         if latest == 'a':
7             return 'b'
8         elif latest == 'b':
9             return 'c'
10        elif latest == 'c':
11            return 'a'
12
13 if RoundRobin.objects.count() == 0:
14     next_flow = 'a'
15 else:
16     latest = RoundRobin.objects.last()
17     next_flow = latest.next()

```

4.4.2 Testing Page Variants

pengujian a/b testing dilakukan dari tanggal 24 September 2017 sampai 23 Oktober 2017. Traffic tinggi terjadi pada tanggal ..., ..., dan ... Yang tertinggi pada tanggal ... dan yang terendah pada tanggal ...

cleaning data membuang dari bot

```

1 <!-- Global Site Tag (gtag.js) - Google Analytics -->
2 <script async src="https://www.googletagmanager.com/gtag/
  js?id=UA-70057602-1">
3 </script>
4 <script>
5   window.dataLayer = window.dataLayer || [];
6   function gtag(){dataLayer.push(arguments);}
7   gtag('js', new Date());
8   gtag('config', 'UA-70057602-1');
9 </script>

```

penerapan data collection dengan google analytic.

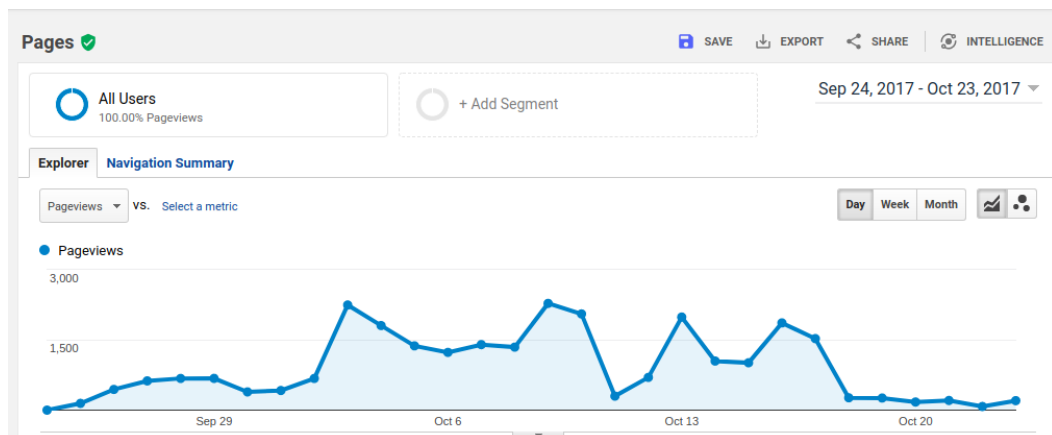


Figure 4.1: Graphic google analytics mekargo.id

menjelaskan gambar.

Primary Dimension: **Page** **Page Title** **Other**

Plot Rows Secondary dimension Sort Type: Default /ukm/ advanced

	Page ?	Pageviews ?	Unique Pageviews ?	Avg. Time on Page ?	Entrances ?	Bounce Rate ?	% Exit ?	Page Value ?
		27,335 % of Total: 97.94% (27,911)	19,711 % of Total: 97.72% (20,170)	00:01:23 Avg for View: 00:01:23 (0.13%)	10,343 % of Total: 98.15% (10,538)	54.08% Avg for View: 53.77% (0.59%)	37.85% Avg for View: 37.76% (0.25%)	\$0.00 % of Total: 0.00% (\$0.00)
1.	/ukm/c/	5,164 (18.89%)	3,801 (19.28%)	00:00:55	3,709 (35.86%)	56.70%	56.66%	\$0.00 (0.00%)
2.	/ukm/b/	4,737 (17.33%)	3,385 (17.17%)	00:01:02	3,281 (31.72%)	52.45%	51.61%	\$0.00 (0.00%)
3.	/ukm/a/	4,222 (15.45%)	3,123 (15.84%)	00:00:52	3,039 (29.38%)	52.45%	50.76%	\$0.00 (0.00%)
4.	/ukm/b/data-diri/	2,200 (8.05%)	1,497 (7.59%)	00:03:30	76 (0.73%)	51.32%	30.32%	\$0.00 (0.00%)
5.	/ukm/c/pinjaman/	1,894 (6.93%)	1,494 (7.58%)	00:01:35	35 (0.34%)	71.43%	38.01%	\$0.00 (0.00%)
6.	/ukm/a/survei/	2,270 (8.30%)	1,392 (7.06%)	00:00:37	44 (0.43%)	54.55%	10.66%	\$0.00 (0.00%)
7.	/ukm/a/detil-survei/	1,913 (7.00%)	1,156 (5.86%)	00:00:31	24 (0.23%)	50.00%	6.33%	\$0.00 (0.00%)
8.	/ukm/a/data-diri/	1,725 (6.31%)	1,127 (5.72%)	00:02:30	22 (0.21%)	59.09%	14.38%	\$0.00 (0.00%)
9.	/ukm/a/pinjaman/	838 (3.07%)	751 (3.81%)	00:02:14	22 (0.21%)	68.18%	30.91%	\$0.00 (0.00%)
10.	/ukm/b/pinjaman/	569 (2.08%)	501 (2.54%)	00:02:21	26 (0.25%)	42.31%	36.20%	\$0.00 (0.00%)

Show rows: 10 Go to: 1 1 - 10 of 21

This report was generated on 1/31/18 at 7:05:00 PM - Refresh Report

Figure 4.2: Table google analytics mekargo.id

4.5 Analyze Test Results

Data yang didapat dari google analytic dilakukan pemrosesan dan analisa dengan metode bayesian inference dengan pymc. Mengimport seluruh library python yang dibutuhkan yaitu pymc3, numpy, figsize, matplotlib, scipy, dan IPython.

```
In [1]: import pymc3 as pm
import numpy as np
%matplotlib inline
from IPython.core.pylabtools import figsize
import matplotlib.pyplot as plt
import scipy.stats as stats
figsize(12.5, 4)
```

Figure 4.3: Import python library

Inisialisasi nilai pada versi A, B, dan C serta banyaknya sample yang digunakan.


```
In [2]: true_A = 120
        true_B = 129
        true_C = 84

        N_sample = 1156
```

Figure 4.4: Initialize sample

Menghitung nilai peluang A, B, dan C. Hasilnya adalah:

```
true_p_A = 0.10380622837370242
```

```
true_p_B = 0.1115916955017301
```

```
true_p_C = 0.0726643598615917
```

```
In [3]: true_p_A = true_A/float(N_sample)
        true_p_B = true_B/float(N_sample)
        true_p_C = true_C/float(N_sample)

        print("true p_A:", true_p_A)
        print("true p_B:", true_p_B)
        print("true p_C:", true_p_C)

true p_A: 0.10380622837370242
true p_B: 0.1115916955017301
true p_C: 0.0726643598615917
```

Figure 4.5: Calculate probability

Melakukan observasi bernoulli terhadap nilai peluang A, B, dan C menggunakan method stats pada library scipy

```
In [4]: observations_A = stats.bernoulli.rvs(true_p_A, size=N_sample)
        observations_B = stats.bernoulli.rvs(true_p_B, size=N_sample)
        observations_C = stats.bernoulli.rvs(true_p_C, size=N_sample)

        print("Obs from Site A: ", observations_A[:30], "...")
        print("Obs from Site B: ", observations_B[:30], "...")
        print("Obs from Site C: ", observations_C[:30], "...")

Obs from Site A: [0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0] ...
Obs from Site B: [0 1 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0] ...
Obs from Site C: [0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0] ...
```

Figure 4.6: Bernoulli observation

Mencetak nilai jumlah dari nilai true yang dihasilkan oleh observasi bernoulli,

dengan hasil sebagai berikut

```
sum_true_p_A = 114
```

```
sum_true_p_B = 138
```

```
sum_true_p_C = 86
```

Mencetak hasil nilai rata-rata observasi bernoulli dengan bantuan method mean pada library numpy. Nilai yang dihasilkan adalah:

```
mean_p_A = 0.098615916955
```

```
mean_p_B = 0.11937716263
```

```
mean_p_C = 0.0743944636678
```

Nilai rata rata ini sekaligus menjadi nilai true frequency

```
In [5]: print(np.mean(observations_A))
        print(np.mean(observations_B))
        print(np.mean(observations_C))

        print(np.sum(observations_A))
        print(np.sum(observations_B))
        print(np.sum(observations_C))

0.0951557093426
0.117647058824
0.0726643598616
110
136
84
```

Figure 4.7: Sum and mean of bernoulli observation

Pymc probability models, jelasin dari model uniform sampling, deterministik nilai posterior a-b, a-c, bc. deterministik bernouli ke model, metropolis dan sampling total data 20000 diambil 1000 awal

```

In [6]: with pm.Model() as model:
        p_A = pm.Uniform("p_A", 0, 1)
        p_B = pm.Uniform("p_B", 0, 1)
        p_C = pm.Uniform("p_C", 0, 1)

        # Define the deterministic delta function. This is our unknown of interest.
        delta_A_B = pm.Deterministic("delta_A_B", p_A - p_B)
        delta_A_C = pm.Deterministic("delta_A_C", p_A - p_C)
        delta_B_C = pm.Deterministic("delta_B_C", p_B - p_C)

        # Set of observations, in this case we have three observation datasets.
        obs_A = pm.Bernoulli("obs_A", p_A, observed=observations_A)
        obs_B = pm.Bernoulli("obs_B", p_B, observed=observations_B)
        obs_C = pm.Bernoulli("obs_C", p_C, observed=observations_C)

        step = pm.Metropolis()
        trace = pm.sample(20000, step=step)
        burned_trace=trace[1000:]

100%|██████████| 20500/20500 [00:31<00:00, 643.38it/s]

```

Figure 4.8: Pymc probabilistic model

menyimpan hasil processing pymc probabilistik model ke variable - variable sebagai bentuk visualisasi dan hasil analisis data

```

In [7]: p_A_samples = burned_trace["p_A"]
        p_B_samples = burned_trace["p_B"]
        p_C_samples = burned_trace["p_C"]
        delta_A_B_samples = burned_trace["delta_A_B"]
        delta_A_C_samples = burned_trace["delta_A_C"]
        delta_B_C_samples = burned_trace["delta_B_C"]

```

Figure 4.9: Pymc probabilistic model results

posterior distribution probabilistik A, deskripsiin isi codingan

```
In [8]: figsize(12.5, 10)

#histogram of posteriors

ax = plt.subplot(311)

plt.xlim(.06, .2)
plt.hist(p_A samples, histtype='stepfilled', bins=25, alpha=0.85,
        label="posterior of $p_A$", color="#A60628", normed=True)
plt.vlines(true_p_A, 0, 80, linestyle="--", label="true $p_A$ (unknown)")
plt.legend(loc="upper right")
plt.title("Posterior distributions of $p_A$, $p_B$, and delta unknowns")
```

Out[8]: Text(0.5,1,'Posterior distributions of \$p_A\$, \$p_B\$, and delta unknowns')

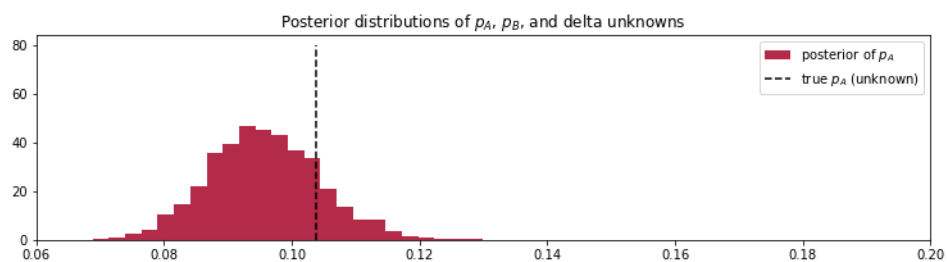


Figure 4.10: Posterior distribution P_A

posterior distribution probabilitas B, deskripsiin isi codingan

```
In [9]: ax = plt.subplot(312)

plt.xlim(.06, .2)
plt.hist(p_B samples, histtype='stepfilled', bins=25, alpha=0.85,
        label="posterior of $p_B$", color="#467821", normed=True)
plt.vlines(true_p_B, 0, 80, linestyle="--", label="true $p_B$ (unknown)")
plt.legend(loc="upper right")
```

Out[9]: <matplotlib.legend.Legend at 0x7f03334969e8>

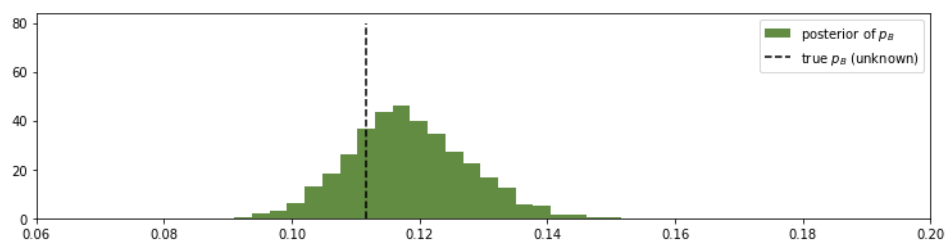


Figure 4.11: Posterior distribution P_B

posterior distribution probabilitas C, deskripsiin isi codingan

```
In [10]: ax = plt.subplot(313)

plt.xlim(0, .1)
plt.hist(p_C_samples, histtype='stepfilled', bins=25, alpha=0.85,
        label="posterior of  $p_C$ ", color="#D6F841", normed=True)
plt.vlines(true_p_C, 0, 80, linestyle="--", label="true  $p_C$  (unknown)")
plt.legend(loc="upper right")
```

Out[10]: <matplotlib.legend.Legend at 0x7f0332445a90>

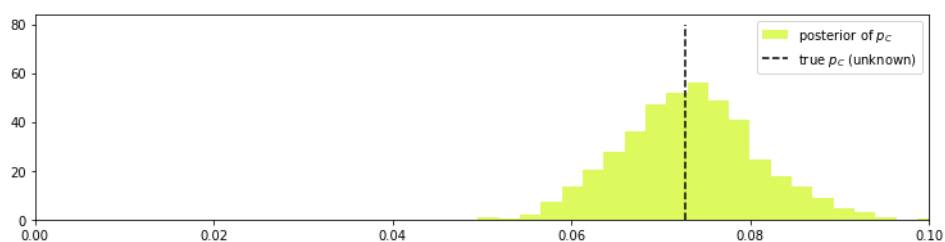


Figure 4.12: Posterior distribution P_C

posterior distribution delta A-B, deskripsiin isi codingan

```
In [11]: ax = plt.subplot(313)
plt.hist(delta_A_B_samples, histtype='stepfilled', bins=30, alpha=0.85,
        label="posterior of delta A-B", color="#7A68A6", normed=True)
plt.vlines(true_p_A - true_p_B, 0, 60, linestyle="--",
        label="true delta (unknown)")
plt.vlines(0, 0, 60, color="black", alpha=0.2)
plt.legend(loc="upper right");
```

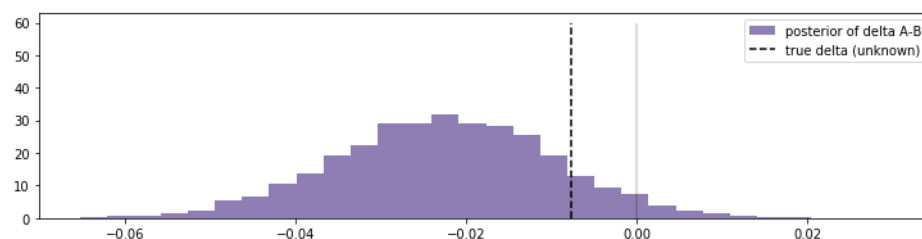


Figure 4.13: Posterior distribution delta A-B

Chapter 5

CONCLUSION AND SUGGESTIONS

5.1 Conclusion

Kesimpulan merupakan jawaban menyeluruh terhadap persoalan yang dipertanyakan di dalam pendahuluan yang merupakan sintesis hasil penelitian & pembahasan. Angka-angka yang dihasilkan dari analisis data (jika penelitian merupakan analisis data) tidak perlu diulang lagi pada bab terakhir ini. Yang diperlukan adalah analisis sebab dihasilkannya angka-angka tersebut.

5.2 Suggestions

Saran dapat berupa saran yang diajukan terhadap persoalan yang diamati dan diteliti, yaitu usul pengembangan atau penyempurnaan penelitian. Saran dapat juga berupa alternatif yang diusulkan pada pihak-pihak terkait.

Bibliography

- Fandango, A. and Idris, I. (2017). *Python Data Analysis - Second Edition*. Packt Publishing.
- Hunter, J., Dale, D., Firing, E., and Droettboom, M. (2014). The matplotlib development team. matplotlib: Python plotting – documentation. 2013.
- Jones, E., Oliphant, T., and Peterson, P. (2014). {SciPy}: open source scientific tools for {Python}.
- Kluyver, T., Ragan-Kelley, B., Pérez, F., Granger, B. E., Bussonnier, M., Frederic, J., Kelley, K., Hamrick, J. B., Grout, J., Corlay, S., et al. (2016). Jupyter notebooks-a publishing format for reproducible computational workflows. In *ELPUB*, pages 87–90.
- Lubanovic, B. (2014). *Introducing Python: Modern Computing in Simple Packages*. O'Reilly Media.
- Lutz, M. (2013). *Learning Python: Powerful Object-Oriented Programming*. Safari Books Online. O'Reilly Media.
- Martin, O. (2016). *Bayesian Analysis with Python*. Packt Publishing.
- Ragan-Kelley, M., Perez, F., Granger, B., Kluyver, T., Ivanov, P., Frederic, J., and Bussonnier, M. (2014). The jupyter/ipynb architecture: a unified view of computational research, from interactive exploration to communication and publication. In *AGU Fall Meeting Abstracts*.
- Van Der Walt, S., Colbert, S., and Varoquaux, G. (2016). The numpy array: a structure for efficient numerical computation (2011).

- Vo, T., Czygan, M., et al. (2015). *Getting started with Python data analysis: learn to use powerful Python libraries for effective data processing and analysis*. Packt Publ.
- Wilson, G., Perez, F., and Norvig, P. (2014). Teaching computing with the ipython notebook. In *Proceedings of the 45th ACM technical symposium on Computer science education*, pages 740–740. ACM.

APPENDIX

CURRICULUM VITAE

IDENTITAS DIRI

Name : Alvian Dwi Kurnianto

Birth Place and Date : Pekalongan, 4 September 1994

Sex : Male

NIM :

Address :

Phone Number :

Email : alviandk@gmail.com

FORMAL EDUCATION

(tahun) – (tahun) :

___ – ___ :

___ – ___ :

WORK EXPERIENCE

(tahun) – (tahun) :

___ – ___ :

___ – ___ :

SCIENCE PUBLICATION (2016)

_____ :

Alvian Dwi Kurnianto, S.Kom.